Anchoring Causal Connections in Physical Concepts

Mario Hubert*, Roland Poellinger†

April 20, 2013

Abstract

In their paper *How Fundamental Physics represents Causality* (2014) Andreas Bartels and Daniel Wohlfarth maintain that there is place for causality in General Relativity. Their argument contains two steps: First they show that there are time-asymmetric models in General Relativity, then they claim to derive that two events are causally connected if and only if there is a time-asymmetric energy flow from one event to the other. In our comment we first give a short summary of their paper followed by a section introducing and pondering different conceptions of causation since Bartels and Wohlfarth avoid declaring explicitly which notion of causation they build on in the paper. In order to analyze their argument in detail we formalize their crucial step in logical terms. This helps to sharpen the question whether their proposed derivation is not just a definition.

1 Grounding (not Reducing) Causality

The question about causality seems to be ubiquitous in the natural sciences and especially in physics. Why and how did the universe evolve? Why is there life on earth? What caused a particle to have that particular trajectory? But if we look closer and examine our physical theories in detail, the issue of how exactly to answer such questions is not that clear. Bertrand Russell was famous for stressing that our fundamental physical theories do not tell us anything about causal relations in their domain since the equations figuring in these theories are time-reversal invariant, and causation seems to be parasitic on time-asymmetry.[1] Even present-day philosophers of science, such as John Norton, do not grant causality any role in physics, because there is no *universal principle of causality that holds true of our science* unless physical theories are *restricted to appropriately hospitable domains.*[2]

---

*University of Lausanne (UNIL), ✉ Mario.Hubert@unil.ch
†MCMP/LMU Munich, ✉ R.Poellinger@lmu.de
[1]Cf. (Russell, 1912).
[2]Cf. (Norton, 2007).

In contrast, physicists themselves normally adhere to an intuition very close to what is known as causal fundamentalism, which claims that the job of all of physics is to uncover the prevalent causal relations in nature. In one of the most famous modern textbooks on classical electrodynamics one finds the following pointed affirmation –

> [. . .] the most sacred tenet in all of physics: the principle of causality.[3]

Neo-Russellians, such as Huw Price, in general still deny causal relations any role in physics, but acknowledge their importance in the special sciences and in common sense reasoning.

However, there are also philosophers like Mathias Frisch, who take up a position between the Neo-Russellians and the causal fundamentalists. Frisch for example claims that causal notions can play a role in physics and do play a role in certain domains when one adds an interpretation to the formal apparatus of a physical theory, which finally allows us to apply the machinery of causal reasoning again.

A seemingly imperative task prior to the quest for causal relations in physics is to try to establish the direction of time within the physical theories under consideration, e.g., with the help of entropy and the second law of thermodynamics (thermodynamic arrow of time), the expansion of the universe (cosmological arrow of time), or the direction of retarded waves (radiative arrow of time). Although these explanations of the direction of time are not unproblematic, the question we are concerned with here is this one: How can we use the time-asymmetry within physical theories in searching for causality?

Following this route, Andreas Bartels and Daniel Wohlfarth set out to show in their paper *How Fundamental Physics represents Causality*[4] that there is place for causality in one of our most successful fundamental theories, namely General Relativity (GR). As time-asymmetry is supposed to be a necessary condition for causal relations, their strategy is first to show in what sense General Relativity is time-asymmetric. Having done that, they provide a description on how to build causality on it. Throughout the paper they emphasize that the time-reversal invariance of the equations does not imply that the solutions must be time-symmetric.

This point is often ignored; so we explicitly want to give the definitions. An *equation* – respectively a law if the law is mathematically formulated as an equation – is said to be time-symmetric (time-reversal invariant) iff for any solution $f(t)$ of that equation $f(-t)$ also presents a solution. We call a *function* $f(t)$ – which is in the most interesting cases a solution of a law-like

---

[3]Cf. (Griffiths, 1999, p. 425).
[4]Cf. (Bartels and Wohlfarth, 2014).

equation – time-symmetric iff there is a t_0 such that $f(t_0 + t) = f(t_0 - t)$ for all t.

The construction of time-asymmetry by Bartels and Wohlfarth relies on the work of Mario Castagnino,[5] who considers models of space-time which can be described by the scale factor and a matter field. It turns out that these models are typically time-asymmetric w.r.t. cosmic time.

The next task is to deduce *local* time-asymmetry by constructing a non-vanishing, continuous, time-like vector field on the time-orientable space-time. Starting from the energy-momentum tensor

$$T_{\mu\nu} = \frac{1}{8\pi} \left(R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - \Lambda g_{\mu\nu} \right), \tag{1}$$

with $R_{\mu\nu}$ the Ricci tensor, R the Ricci curvature, Λ the cosmological constant, and $g_{\mu\nu}$ the metrical tensor, we get (after imposing further restrictions on the energy-momentum tensor)

$$T_{\mu\nu} = s_0 V_\mu^0 V_\nu^0 + \sum_{i=1}^{3} s_i V_\mu^i V_\nu^i, \tag{2}$$

such that $s_0 \geq 0$, $s_i \in \mathbb{R}$, V_ν^i are space-like vector fields for all $i \in \{1, 2, 3\}$, and V_μ^0 is a time-like vector field on the space-time manifold. Making explicit the space-time coordinate x, $V_\mu^0(x)$ is the supposed non-vanishing, continuous, time-like vector field that establishes the necessary local time-asymmetry by being a means of distinguishing between the semi-light-cones on every space-time point. Furthermore, there is a physical quantity connected to $V_\mu^0(x)$: T_μ^0 is interpreted as the energy flow in the direction of $V_\mu^0(x)$.

Finally, causality may enter the arena as stated in the paper:

> [W]e find a physical vector field on which the *time-asymmetric causal connection* between events can be based. We will say that events C and E are *causally connected* iff there is a time-asymmetric energy flow from C to E.

This is the crux of how Bartels and Wohlfarth anchor causality in General Relativity. We will discuss their proposal in the following. In Section 2 we introduce the most prominent theories of causation and ask what concept of causation Bartels and Wohlfarth use. We then examine in Section 3 the details of how they ground causality in General Relativity upon formalizing the above quotation. Section 4 presents a short summary and outlook.

---

[5]Cf. (Castagnino et al., 2003a,c,b; Castagnino and Lombardi, 2009).

2 Concepts of Causation

Outside of physics (many times on the border), philosophers have been thinking systematically about cause and effect since the very beginnings and even more rigorously with the introduction of mathematical methods and formal semantics into the discipline in the last century. Epistemology and philosophy of science at once had the means to shape prevailing problems in symbolic form, express achievements with scientific stringency, and sort issues within formal theories from questions about intuitions and basal premisses. Select approaches to characterizing causal relations shall be outlined in the following to give an overview over the problems a causal theorist is facing when casting plausible analysis in formal structure.

David Hume's famous quote may be seen as the point of reference for many formal theories – he makes out an essentially structural unifying feature of causal relations when he claims in 1748 that *"[w]e may define a cause to be an object followed by another, and where all objects, similar to the first, are followed by objects similar to the second."* This portion of *An Enquiry about Human Understanding* (Section VII) becomes the corner stone of the regular and ultimately probabilistic analysis of causation connected with the names of, e.g., Good and Reichenbach or also Suppes, who explicitly builds the direction of time into his account to express our intuitions about the *temporal* asymmetry in formal manner: A cause must necessarily be correlated with its effect *and* precede it.

Nancy Cartwright is well-known for her critique of a purely mathematical, *thin* characterization of causation, especially of the one based merely on regularities or correlations – she ultimately emphasizes the experimenter's knowledge about the experimental setup and ties methodology and interpretation together as two sides of the same causal coin. In her eyes, transferring causal knowledge from narrowly defined lab conditions to situations of larger scale or everyday experience cannot follow one single principle, on the contrary, it must naturally be as diverse (though maybe family-like) in character as the methodology applied in the first place itself. What is important for our discussion, though, is her standpoint that causality cannot be a monolithic, fundamental concept due to its multifarious nature.

The proponents of an interventionist accounts of causation agree with Cartwright on one central thing: They acknowledge the problems of a mere statistical analysis of causality. The interventionist theorist wants to do better by adding an element of counterfactual analysis to the probabilistic framework without being accused at the same time of metaphysically overloading our mostly solid notion of causation. Structural hypothetical interventions, utilized for the test of causal efficacy and formally expressed as fixing the value of a (random) variable in a Bayesian network, connect with a scientist's practice and mathematical toolbox at the same time – without causation coming under the suspicion of being reduced to an an-

thropomorphic concept. Causes are expressed as difference-makers in given circumstances.

Interventionist theorists as Judea Pearl clearly localize causal relationships with underlying physical mechanisms on the objective side of things and dismiss a subjective or even epistemic reading. The interventionist framework is open to this reading, nevertheless. Ultimately, deciding upon the set of variables considered illuminating for the analysis to be conducted is obviously a subjective (sometimes highly pragmatic) process that may differ from one epistemic agent to the next even if performed in compliance with rational standards. Jon Williamson, for example, is opting for a fruitful exploration of causal graphs as bearers of epistemic contents and direct enablers of meaningful communication. Cases of causation by omission, the distinction whether a result was actively obtained or passively not prevented, as well as the question how to determine accountability on the basis of causal efficacy can be made transparent in an epistemic account without much hassle.

Now, when Bartels and Wohlfarth set out to search for fundamental prints of causation in physical grounds, which of the many readings of causation do they have in mind, or in other words: What might constitute a good candidate set of features for them that makes a causal relation causal?

Just as Bartels and Wohlfarth some philosophers have tried to take a different perspective and approach the task of formalizing intuitions about causality from a point of view closer to physics. The causal powers theorist is straightforwardly asking the question why not to introduce causality as a basic power and ascribe essential causal capacities to objects of reality. *Dispositions* are meant to be necessarily separate from their token instantiations, but at the same time linked to those instantiations of themselves through a necessary causal relation: Causal powers (as Popper's *propensities*) are seen as enduring states with the disposition to objectively produce events or states by singularly contributing observable quantities to their manifestations. Problems with this account arise as soon as we ask about the nature of the connection between those powers and their manifestations. And: Can one really postulate a certain disposition if it, for example, never manifests itself? And if we want time-asymmetry to be built into the expression of a causal relation: Is there a way to understand the directedness of powers as necessary causal directedness from cause towards effect?

Processes seem to be another promising fundamental building block candidate for a theory of causation. At the core of process theories lies the explication of world lines and their intersection, understood as more basic than the causal relation between events. Phil Dowe extends Wesley Salmon's material work[6] by introducing exchanged conserved quantities, such as linear momentum, mass-energy, or charge, to make it empirically applicable.

---

[6]Cf. (Salmon, 1997).

Dowe's theory relies on the following two propositions:[7]

1. A *causal interaction* is an intersection of world lines which involves exchange of a conserved quantity.

2. A *causal process* is a world line of an object which possesses a conserved quantity.

A *world line* is the collection of points in space-time that represents the history of an object. A *process* is understood as the world line of an object, regardless of whether or not it possesses any conserved quantities. As we are here concerned with causation in fundamental physics, an *object* is anything found in the ontology of a fundamental physical theory, e.g., particles, waves, or fields. A *conserved quantity* is any quantity that is universally conserved in our actual physical world. Our current fundamental physical theories tell us what these quantities are (most prominently energy) and by which laws they are governed.

An *intersection* simply is the overlapping of two or more processes in space-time – it consists of all space-time points common to both (or all) processes. An *exchange* occurs when at least one incoming and at least one outgoing process undergo a change in the value of the conserved quantity – in this case, "incoming" and "outgoing" are defined with respect to the light-cone structure of space-time.

With these definitions at hand one can finally state that events C and E are causally connected (connected by a causal relation) iff a continuous series of causal processes and mediating interactions can be traced between them. In this very coarse first draft of a naïve process theory, however, it seems that routine analysis returns too many causes, which can only be reduced again by utilizing extra-theoretical knowledge or assumptions (against the original goal of theoretically objectivizing the notion of a causal process).

Moreover, although the conserved quantity theory seeks to ground causality in physics it seems not to be compatible with GR.[8] The most important reason for this is the lack of an energy conservation law for most models of GR. In general, energy is not a conserved quantity in this theory, and for this very reason the notion of a physical process becomes meaningless if it is to be built upon any definition of energy.

One can of course reply that the conserved quantity theory is applicable to isolated physical systems where it is possible to reclaim conservation of energy, and restrict oneself to these cases. However, isolated systems are idealizations that are an abstraction of our actual world and, strictly speaking, cannot be found therein. Current research raises further questions along

---

[7]Cf. e.g. (Dowe, 2000).

[8]The arguments for this proposition are taken from (Lam, 2005) referring to (Curiel, 2000) and (Rueger, 1998).

these lines and beyond our considerations in this paper: Is it possible that our actual world can perhaps be described by some space-time that allows for a conservation of energy within GR? We shall move on to Bartels and Wohlfarth's bridge building between causal claims and physical terms.

3   Bridging Causality and Energy Flow

In order to elicit the energy-flow T^0_μ from GR, Bartels and Wohlfarth impose a myriad of restrictions on the space-time model they use. They only consider models that can be described by the scale factor plus a matter field. Almost all of these models turn out to be time-asymmetric w.r.t. global time. To establish local time-asymmetry and consequently energy-flow they introduce some further – rather technical – constraints on the energy-momentum tensor. Is the resulting space-time model a model of the universe we live in? Is causality thus grounded in the physical world or just in a special solution of GR? Is it also possible to ground causality in space-time models other than the ones described by a scale factor?

Bartels and Wohlfarth's central statement finally is the formulation of the link between two events' causal and physical relation. Their ultimate goal is to associate the time-asymmetric causal connection between events with physical concepts in the aforementioned claim that

"events C and E are *causally connected* iff there is a time-asymmetric energy flow from C to E."

This statement summarizes Bartels and Wohlfarth's view that causal asymmetry is not to be conventionally defined by concepts anchored in physics, but that it is rather "derived from global time-asymmetry" and thus intrinsically physically endowed with two distinct directions.

We read the formulation "events C and E are causally connected" as "event C causally contributes to event E" in order to avoid the interpretation of C as the true, the sole, or the actual cause of E and, at the same time, to avoid talk of prevented or potential causation and the like. To us *causally connected* is just a very weak notion conveying that the event C plays a certain (yet to be determined) causal role for the occurrence of E. However, Bartels and Wohlfarth neither explain what they mean by *causal connection* nor state what notion of causality they intend to ground in physics. Instead, they introduce the following restrictions:

> In order to avoid misunderstanding, we want to make it clear in advance that we do not aim at a *reduction* of the concept of causation to physics, i.e. we do not propose that causation can in general be *defined* in terms of fundamental physical relations as proposed by transfer theories of causation. [...] We would also not propose that notorious problems like the problem of absences and non-occurrences as causes can be sufficiently tackled

by means of a transfer theory of causation. Counterfactual models of causal reasoning may well play their role in physics, but we see no way to generally define the notions employed in counterfactual models of causation by recourse to elementary physical relations (such as energy-momentum transfer).

Clearly, Bartels and Wohlfarth do not claim that all causal relations can be explained as relations described by fundamental physics, nor do they want to include omissions or counterfactual statements in their examination. They continue:

> We are exclusively concerned with causal asymmetry as a *sine qua non* condition for the existence of causal relations – however those relations may further be conceptualized in order to answer causal questions occurring in particular contexts.

Causal reasoning in physical contexts fundamentally relies on causal asymmetry: If C causes E, then E does not cause C. Coupling this causal asymmetry with temporal asymmetry we get that C is temporally prior to E if C causes E. It is this line of reasoning, which seems to be Bartels and Wohlfarth's motivation for grounding causal relations in the time-asymmetry of certain models of GR.

Now, in a first attempt to carve out the logical structure of Bartels and Wohlfarth's central claim we might look at the following formula:

$$\forall C, E \in \mathcal{E} \left(C \twoheadrightarrow E \iff \exists f (C \overset{f}{\rightsquigarrow} E) \right) \tag{3}$$

with $\mathcal{E}$ the global set of events, '$\twoheadrightarrow$' representing the time-asymmetric causal connectedness between two events, and '$\overset{f}{\rightsquigarrow}$' representing the time-asymmetric energy flow f between two events.

Now, as far as we can see all causal theories agree with '$\Longleftarrow$', in fact, most (if not all) formal theories of causation will use this direction as one of the crucial applicability benchmarks – in general: The postulation of energy flow between two events supports the intuition that these two events are bound causally in a push-pull way, even if C is not an actual, maybe only a contributing cause of some effect E.

The other direction, though, seems to raise some questions: The set $\mathcal{E}$ will have to be restricted in a suitable manner to allow for such an inference. As said above, omissions and non-occurrences are excluded by Bartels and Wohlfarth themselves, who admittedly aim for some (limited) concept of *physical causation*.

Now, any potential restrictions should be explicitly expressible – we might consequently modify our formula above by imposing a set of conditions Γ on C and E:

$$\forall C, E \in \mathcal{E} \left(\Gamma \implies \left(C \rightarrow E \iff \exists f (C \overset{f}{\rightsquigarrow} E) \right) \right) \tag{4}$$

What might Γ stand for? If the premises of the right-hand side (e.g., the existence of a distinguished physical vector along which the direction of energy flow is to be aligned) turn out to constitute the basis for the left-hand side as well, the biconditional thus formulated might become insubstantially *thin* in the end. If by narrowing down potential C–E pairs only process-like connected Cs and Es remain, we are essentially left with postulating 'measurable exchange of conserved quantities' as an explication of '$\rightarrow$', or simply put: 'energy flow'. The question remains: If the proposed biconditional is not a definition, what intuitions ought '$\rightarrow$' to capture above and beyond '$\exists f (\cdot \overset{f}{\rightsquigarrow} \cdot)$'?

4   Arrows and Targets

Bartels and Wohlfarth's paper provides an interesting answer to Russell's fundamental critique in considering time-asymmetric solutions of a time-symmetric physical law on which causal asymmetry might then be based in turn, "a *sine qua non* condition for the existence of causal relations", in any case a condition that usually has the status of an unquestioned, almost axiomatic precondition for causal analysis. The authors consequently aim to establish a *weak* causal arrow:

> Causal relations between events have a substantial (not conventional) time-direction that is in line with one of the global time-directions which are substantially (not conventionally) different in virtue of their particular geometrical characteristics.

In their investigations the authors make out the asymmetry of global time upon which causal relations might be based, finally. However, this should be distinguished from a much stronger claim: the problem of the *strong* causal arrow, where one has to show additionally which global direction is the future and which the past direction we experience in daily life. Using this distinction one has to be attentive not to mix problems concerning the arrow of time and causality, though they are surely associated, as we could see in this paper.

We agree with Bartels and Wohlfarth's critique of (Castagnino and Lombardi, 2009) at the end of their paper: It is too simple to merely stipulate that the future direction coincides with the direction of *positive* local energy flow. Instead, one should state that the geometrical development of space with respect to global time characterizes future and past. This can then be utilized to define local future and local past and to finally anchor causal connections in physical concepts – in models where the transition from

9

global time to a local one is possible. However, although the directionality of time might well be the common thread of asymmetric energy flow and the asymmetry of causation, it still remains unclear in the paper why exactly '$\rightarrow$' should be named 'causal', and if so, what exactly the meaning of this attribute (undefined until last) might actually be.

Looking at the big picture, it will be inevitable to determine whether our actual world can be described in accordance with the physical restrictions Bartels and Wohlfarth impose on the general relativistic space-time model they use in the first place, such that their causal notion can truly be based in nature. We find the idea of geometrically grounding causation in physics by formally building on time-asymmetry very fruitful – next targets could be the problem of the strong causal arrow as well as the clarification of how this construction connects with other theoretical approaches towards causation.

References

Bartels, A. and Wohlfarth, D. (2014). How fundamental physics represents causality. In Galavotti, M. C. et al., editors, *New Directions in the Philosophy of Science*. Springer.

Castagnino, M., Lara, L., and Lombardi, O. (2003a). The cosmological origin of time asymmetry. *Classical and Quantum Gravity*, 20(2):369.

Castagnino, M., Lara, L., and Lombardi, O. (2003b). The direction of time: From the global arrow to the local arrow. *International Journal of Theoretical Physics*, 42:2487–2504.

Castagnino, M. and Lombardi, O. (2009). The global non-entropic arrow of time: From global geometrical asymmetry to local energy flow. *Synthese*, 169(1):1–25.

Castagnino, M., Lombardi, O., and Lara, L. (2003c). The global arrow of time as a geometrical property of the universe. *Foundations of Physics*, 33:877–912.

Curiel, E. (2000). The constraints general relativity places on physicalist accounts of causality. *Theoria – Segunda Época*, 15(1):33–58.

Dowe, P. (2000). *Physical Causation*. Cambridge: Cambridge University Press.

Griffiths, D. (1999). *Introduction to electrodynamics*. Upper Saddle River, New Jersey: Prentice Hall.

Lam, V. (2005). Causation and space-time. *History and Philosophy of the Life Sciences*, 27(3–4):465–478.

Norton, J. D. (2007). Causation as folk science. In Price, H. and Corry, R., editors, *Causation, Physics and the Constitution of Reality*. New York: Oxford University Press.

Rueger, A. (1998). Local theories of causation and the a posteriori identification of the causal relation. *Erkenntnis*, 48:25–38.

Russell, B. (1912). On the notion of cause. In *Proceedings of the Aristotelian society*, volume 13, pages 1–26.

Salmon, W. (1997). Causality and explanation: A reply to two critiques. *Philosophy of Science*, 64:461–77.